Amendments to the Specification

Replace the paragraph starting on page 1, line 20 with the following:

"A NOVEL APPROACH TO SPEECH RECOGNITION", U.S. Patent Application Serial Number 09/815,768~~, attorney docket number ELZK-001~~;

"REMOTE SERVER OBJECT ARCHITECTURE FOR SPEECH RECOGNITION", U.S. Patent Application Serial Number 09/815,808~~, attorney docket number ELZK-003~~; and

"WEB-BASED SPEECH RECOGNITION WITH SCRIPTING AND SEMANTIC OBJECTS ", U.S. Patent Application Serial Number 09/815,726~~, attorney docket number ELZK-004~~.

Replace the paragraph starting on page 1, line 20 with the following:

The functionality of the present invention may be hosted on a single device or distributed in any of a variety of manners among several devices. Such devices may be networked together or accessible over any of a variety of networks such as the Internet, World Wide Web ("Web"), intranet, extranet, local area network (LAN), wide area network (WAN), private network, virtual network, virtual private network (VPN), telephone network, cellular telephone network, cable network, or some combination thereof, as examples. When implemented in a Web setting, the present invention may be implemented using Web-based technologies, such as by scripting a transactional application system within the context of a Web page, as described in co-pending U.S. Patent application Serial Number 09/815,726 ~~(Attorney's references ELZK-004)~~,

incorporated herein by reference.

Replace the paragraph starting on page 8, line 6 with the following:

The syntactic description includes a list of alternatives or sequences. Each sequence may be a list of items, where the items may be either words or other class instances. Each class also has an optional semantic description that includes a list of semantic attributes. Each semantic attribute may be a value, a category, an operator, or a tree of such things. Attribute values are specific items, such as the number 3, that have meaning when interpreted at run-time. Categories are symbols, possibly with values, that mark the path for future semantic interpretation. Operators control the combination of class instances and provide a powerful, extensible, and general technique for semantic evaluation. Note that any given class may have, and may be interpreted in accordance with, multiple categories. These categories control different semantic interpretations of the same class instance. Collectively, the categories describe all possible valid interpretations of the class. Because all classes are context free, they may be re-used and reinterpreted in accordance with different contexts. For example, a class representing the numbers from 20 – 99 may be reused in several instances where there is a phonetic input corresponding to a number.

Replace the paragraph starting on page 8, line 20 with the following:

A phonetic search module is configured to receive input phonetic data and generate a corresponding best word list (including word paths) using

syntactic analysis. Typically, the input phonetic data will be in response to a prompt. The prompt is provided (e.g., a question) is asked) within a given context, so a response within a certain realm of responses is expected according to that context. The phonetic search module includes a phonetic search algorithm (PSA) used to search the CFG DB for the best matching words and/or phrases corresponding to the phonetic stream input and the grammars associated with the context. The PSA is a two-layer search algorithm, working like a phrase spotter with skips. The first layer converts the incoming phonetic data, comprised of sonotypes, into sequences of words, generating only the ones allowed by currently active grammars. A sonotype is a phonetic estimate of a syllable or word.

Replace the paragraph starting on page 9, line 9 with the following:

To "spot" words from the received phonetic stream, the phonetic search module applies word models to the sonotypes and score restrictions. In the present invention, each sonotype is represented along a timeline as having a first portion that represents the phonetic information starting at a start time and then concluding at a first end time and having a first score. A second portion begins at the first end time and ends at a final end time and includes additional phonetic information derived from the original audio input. Therefore, unlike prior systems, the end times for each sonotype are not fixed. Each end time corresponds to a point in time that the speaker may have finished uttering the given sonotype, and will yield a different score representing the probability that

the utterance was a certain word. For example, the word "yes" may be modeled and include one start time and six different end times, each end time having a different score associated therewith. Applying the word model to the first group of minimum phonetic information (i.e., from start time to the first end time), yields a word (or syllable) result with a certain score. Applying the word model to a second group of phonetic information (i.e., from start time to a second, later end time) yields a different score. Using this modeling, a set of words is determined.

Replace the paragraph starting on page 10, line 2 with the following:

While the first layer of the PSA generates words, the second layer of the PSA search algorithm includes a grammar builder that connects consecutive words, represented as segments, into grammar instances that define word paths. For example, a word path may be (start) yes-I-do (end), where each word is a sonotype. The word "yes" may be the first word segment in a path and the words "I" and "do" may follow as subsequent segments. The process of connecting word segments into phrases is accomplished as a further function of the word representations mentioned above, with a plurality of possible end times. In accordance with the rules implemented in the present invention, a first word segment can only be connected with a second word segment if the second word begins after conclusion of the first word. Given the possibility of multiple end times for a given word representation, the second word may start after a first (i.e., earlier) end time and prior to a second (i.e., later) end time. In that

case, a connection between those word segments can exist. In the case where the second word begins prior to the first end time, a connection can not exist. By making these connections and combining segments, word paths are formed by the grammar builder. The output of the grammar builder is referred to as a best word list, which includes the words and paths, referred to as sequences. That is, for a given word list, many word paths and sequences of those words may be possible.